

## **RAW SEQUENCE LISTING**

**The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) no errors detected.**

Application Serial Number: 09/917,384B

Source: FW/6

Date Processed by STIC: 9/9/05

# ***ENTERED***



IFW16

## RAW SEQUENCE LISTING

DATE: 09/09/2005

PATENT APPLICATION: US/09/917,384B

TIME: 09:01:38

Input Set : A:\Sequence listing.txt

Output Set: N:\CRF4\09092005\I917384B.raw

```

3 <110> APPLICANT: ADNEY, WILLIAM S
4     DING, SHI-YOU S
5     VINZANT, TODD B.
6     DECKER, STEPHEN R.
7     HIMMEL, MICHAEL E.
9 <120> TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
CELLULOLYTICUS
11 <130> FILE REFERENCE: NREL 01-38
13 <140> CURRENT APPLICATION NUMBER: US 09/917,384B
14 <141> CURRENT FILING DATE: 2001-07-28
16 <160> NUMBER OF SEQ ID NOS: 11
18 <170> SOFTWARE: PatentIn version 3.2
20 <210> SEQ ID NO: 1
21 <211> LENGTH: 1121
22 <212> TYPE: PRT
23 <213> ORGANISM: Acidothermus cellulolyticus
26 <220> FEATURE:
27 <221> NAME/KEY: misc feature
28 <223> OTHER INFORMATION: Full-length sequence of Gux1 protein
30 <400> SEQUENCE: 1
32 Met Pro Gly Leu Arg Arg Arg Leu Arg Ala Gly Ile Val Ser Ala Ala
33 1          5          10          15
36 Ala Leu Gly Ser Leu Val Ser Gly Leu Val Ala Val Ala Pro Val Ala
37          20          25          30
40 His Ala Ala Val Thr Leu Lys Ala Gln Tyr Lys Asn Asn Asp Ser Ala
41          35          40          45
44 Pro Ser Asp Asn Gln Ile Lys Pro Gly Leu Gln Leu Val Asn Thr Gly
45          50          55          60
48 Ser Ser Ser Val Asp Leu Ser Thr Val Thr Val Arg Tyr Trp Phe Thr
49 65          70          75          80
52 Arg Asp Gly Gly Ser Ser Thr Leu Val Tyr Asn Cys Asp Trp Ala Ala
53          85          90          95
56 Met Gly Cys Gly Asn Ile Arg Ala Ser Phe Gly Ser Val Asn Pro Ala
57          100         105         110
60 Thr Pro Thr Ala Asp Thr Tyr Leu Gln Leu Ser Phe Thr Gly Gly Thr
61          115         120         125
64 Leu Ala Ala Gly Gly Ser Thr Gly Glu Ile Gln Asn Arg Val Asn Lys
65          130         135         140
68 Ser Asp Trp Ser Asn Phe Asp Glu Thr Asn Asp Tyr Ser Tyr Gly Thr
69 145         150         155         160
72 Asn Thr Thr Phe Gln Asp Trp Thr Lys Val Thr Val Tyr Val Asn Gly
73          165         170         175
76 Val Leu Val Trp Gly Thr Glu Pro Ser Gly Ala Thr Ala Ser Pro Ser
77          180         185         190

```

## RAW SEQUENCE LISTING

DATE: 09/09/2005

PATENT APPLICATION: US/09/917,384B

TIME: 09:01:38

Input Set : A:\Sequence listing.txt

Output Set: N:\CRF4\09092005\I917384B.raw

```

80 Ala Ser Ala Thr Pro Ser Pro Ser Ser Ser Pro Thr Thr Ser Pro Ser
81      195      200      205
84 Ser Ser Pro Ser Pro Ser Ser Ser Pro Thr Pro Thr Pro Ser Ser Ser
85      210      215      220
88 Ser Pro Pro Pro Ser Ser Asn Asp Pro Tyr Ile Gln Arg Phe Leu Thr
89 225      230      235      240
92 Met Tyr Asn Lys Ile His Asp Pro Ala Asn Gly Tyr Phe Ser Pro Gln
93      245      250      255
96 Gly Ile Pro Tyr His Ser Val Glu Thr Leu Ile Val Glu Ala Pro Asp
97      260      265      270
100 Tyr Gly His Glu Thr Thr Ser Glu Ala Tyr Ser Phe Trp Leu Trp Leu
101      275      280      285
104 Glu Ala Thr Tyr Gly Ala Val Thr Gly Asn Trp Thr Pro Phe Asn Asn
105      290      295      300
108 Ala Trp Thr Thr Met Glu Thr Tyr Met Ile Pro Gln His Ala Asp Gln
109 305      310      315      320
112 Pro Asn Asn Ala Ser Tyr Asn Pro Asn Ser Pro Ala Ser Tyr Ala Pro
113      325      330      335
116 Glu Glu Pro Leu Pro Ser Met Tyr Pro Val Ala Ile Asp Ser Ser Val
117      340      345      350
120 Pro Val Gly His Asp Pro Leu Ala Ala Glu Leu Gln Ser Thr Tyr Gly
121      355      360      365
124 Thr Pro Asp Ile Tyr Gly Met His Trp Leu Ala Asp Val Asp Asn Ile
125      370      375      380
128 Tyr Gly Tyr Gly Asp Ser Pro Gly Gly Gly Cys Glu Leu Gly Pro Ser
129 385      390      395      400
132 Ala Lys Gly Val Ser Tyr Ile Asn Thr Phe Gln Arg Gly Ser Gln Glu
133      405      410      415
136 Ser Val Trp Glu Thr Val Thr Gln Pro Thr Cys Asp Asn Gly Lys Tyr
137      420      425      430
140 Gly Gly Ala His Gly Tyr Val Asp Leu Phe Ile Gln Gly Ser Thr Pro
141      435      440      445
144 Pro Gln Trp Lys Tyr Thr Asp Ala Pro Asp Ala Asp Ala Arg Ala Val
145      450      455      460
148 Gln Ala Ala Tyr Trp Ala Tyr Thr Trp Ala Ser Ala Gln Gly Lys Ala
149 465      470      475      480
152 Ser Ala Ile Ala Pro Thr Ile Ala Lys Ala Ser Gln Thr Gly Asp Tyr
153      485      490      495
156 Leu Arg Tyr Ser Leu Phe Asp Lys Tyr Phe Lys Gln Val Gly Asn Cys
157      500      505      510
160 Tyr Pro Ala Ser Ser Cys Pro Gly Ala Thr Gly Arg Gln Ser Glu Thr
161      515      520      525
164 Tyr Leu Ile Gly Trp Tyr Tyr Ala Trp Gly Gly Ser Ser Gln Gly Trp
165      530      535      540
168 Ala Trp Arg Ile Gly Asp Gly Ala Ala His Phe Gly Tyr Gln Asn Pro
169 545      550      555      560
172 Leu Ala Ala Trp Ala Met Ser Asn Val Thr Pro Leu Ile Pro Leu Ser
173      565      570      575
176 Pro Thr Ala Lys Ser Asp Trp Ala Ala Ser Leu Gln Arg Gln Leu Glu

```

## RAW SEQUENCE LISTING

DATE: 09/09/2005

PATENT APPLICATION: US/09/917,384B

TIME: 09:01:38

Input Set : A:\Sequence listing.txt

Output Set: N:\CRF4\09092005\I917384B.raw

```

177          580          585          590
180 Phe Tyr Gln Trp Leu Gln Ser Ala Glu Gly Ala Ile Ala Gly Gly Ala
181          595          600          605
184 Thr Asn Ser Trp Asn Gly Asn Tyr Gly Thr Pro Pro Ala Gly Asp Ser
185          610          615          620
188 Thr Phe Tyr Gly Met Ala Tyr Asp Trp Glu Pro Val Tyr His Asp Pro
189 625          630          635          640
192 Pro Ser Asn Asn Trp Phe Gly Phe Gln Ala Trp Ser Met Glu Arg Val
193          645          650          655
196 Ala Glu Tyr Tyr Tyr Val Thr Gly Asp Pro Lys Ala Lys Ala Leu Leu
197          660          665          670
200 Asp Lys Trp Val Ala Trp Val Lys Pro Asn Val Thr Thr Gly Ala Ser
201          675          680          685
204 Trp Ser Ile Pro Ser Asn Leu Ser Trp Ser Gly Gln Pro Asp Thr Trp
205          690          695          700
208 Asn Pro Ser Asn Pro Gly Thr Asn Ala Asn Leu His Val Thr Ile Thr
209 705          710          715          720
212 Ser Ser Gly Gln Asp Val Gly Val Ala Ala Ala Leu Ala Lys Thr Leu
213          725          730          735
216 Glu Tyr Tyr Ala Ala Lys Ser Gly Asp Thr Ala Ser Arg Asp Leu Ala
217          740          745          750
220 Lys Gly Leu Leu Asp Ser Met Trp Asn Asn Asp Gln Asp Ser Leu Gly
221          755          760          765
224 Val Ser Thr Pro Glu Thr Arg Thr Asp Tyr Ser Arg Phe Thr Gln Val
225          770          775          780
228 Tyr Asp Pro Thr Thr Gly Asp Gly Leu Tyr Ile Pro Ser Gly Trp Thr
229 785          790          795          800
232 Gly Thr Met Pro Asn Gly Asp Gln Ile Lys Pro Gly Ala Thr Phe Leu
233          805          810          815
236 Ser Ile Arg Ser Trp Tyr Thr Lys Asp Pro Gln Trp Ser Lys Val Gln
237          820          825          830
240 Ala Tyr Leu Asn Gly Gly Pro Ala Pro Thr Phe Asn Tyr His Arg Phe
241          835          840          845
244 Trp Ala Glu Ser Asp Phe Ala Met Ala Asn Ala Asp Phe Gly Met Leu
245          850          855          860
248 Phe Pro Ser Gly Ser Pro Ser Pro Thr Pro Ser Pro Thr Pro Thr Ser
249 865          870          875          880
252 Ser Pro Ser Pro Thr Pro Ser Ser Ser Pro Thr Pro Ser Pro Ser Pro
253          885          890          895
256 Ser Pro Thr Gly Asp Thr Thr Pro Pro Ser Val Pro Thr Gly Leu Gln
257          900          905          910
260 Val Thr Gly Thr Thr Thr Ser Ser Val Ser Leu Ser Trp Thr Ala Ser
261          915          920          925
264 Thr Asp Asn Val Gly Val Ala His Tyr Asn Val Tyr Arg Asn Gly Thr
265          930          935          940
268 Leu Val Gly Gln Pro Thr Ala Thr Ser Phe Thr Asp Thr Gly Leu Ala
269 945          950          955          960
272 Ala Gly Thr Ser Tyr Thr Tyr Thr Val Ala Ala Val Asp Ala Ala Gly
273          965          970          975

```

## RAW SEQUENCE LISTING

DATE: 09/09/2005

PATENT APPLICATION: US/09/917,384B

TIME: 09:01:38

Input Set : A:\Sequence listing.txt

Output Set: N:\CRF4\09092005\I917384B.raw

```

276 Asn Thr Ser Ala Gln Ser Phe Ala Gly Asp Ser Asp Asp Gly Ile Ala
277          980          985          990
280 Val Ala Ser Pro Ser Pro Ser Pro Thr Pro Thr Ser Ser Pro Ser Pro
281          995          1000          1005
284 Thr Pro Ser Pro Thr Pro Ser Pro Thr Ser Thr Ser Gly Ala Ser
285      1010          1015          1020
288 Cys Thr Ala Thr Tyr Val Val Asn Ser Asp Trp Gly Ser Gly Phe
289      1025          1030          1035
292 Thr Thr Thr Val Thr Val Thr Asn Thr Gly Thr Arg Ala Thr Ser
293      1040          1045          1050
296 Gly Trp Thr Val Thr Trp Ser Phe Ala Gly Asn Gln Thr Val Thr
297      1055          1060          1065
300 Asn Tyr Trp Asn Thr Ala Leu Thr Gln Ser Gly Lys Ser Val Thr
301      1070          1075          1080
304 Ala Lys Asn Leu Ser Tyr Asn Asn Val Ile Gln Pro Gly Gln Ser
305      1085          1090          1095
308 Thr Thr Phe Gly Phe Asn Gly Ser Tyr Ser Gly Thr Asn Thr Ala
309      1100          1105          1110
312 Pro Thr Leu Ser Cys Thr Ala Ser
313      1115          1120
316 <210> SEQ ID NO: 2
317 <211> LENGTH: 3365
318 <212> TYPE: DNA
319 <213> ORGANISM: Acidothermus cellulolyticus
322 <220> FEATURE:
323 <221> NAME/KEY: misc_feature
324 <223> OTHER INFORMATION: Gux1 full-length coding sequence
326 <400> SEQUENCE: 2
327 atgccaggat tacgacggcg actccgcgcc ggtatcgtct cggcggcgcc gttgggggtcg      60
329 ctgggttagcg ggctcgttgc cgtcgcacca gtcgcgcacg cggcgggtgac tctcaaagcg      120
331 cagtataaga acaatgattc ggcgccgagt gacaaccaga tcaaaccggg tctccagttg      180
333 gtgaataccg ggctcgtcgtc ggtggatttg tcgacgggtga cgggtgcggta ctgggttcacc      240
335 cgggatgggtg ggctcgtcgac actgggtgtac aactgtgact gggcggcgat ggggtgtggg      300
337 aatatccgcg cctcgttcgg ctcggtgaac cggcgacgc cgacggcgga cacctacctg      360
339 cagttgtcgt tcaactgggtg aacgttggcc gctgggtggg cgacgggtga gattcaaaac      420
341 cgggtgaata agagtgactg gtcgaacttt gatgagacca atgactactc gtatgggacg      480
343 aacaccacct tccaggactg gacgaagggt acggtgtacg tcaacggcgt gttggtctgg      540
345 gggaccgaac cgtccggagc gacggcgtct ccatccgcgt cggcgacgcc cagcccgtcc      600
347 agttcaccga ccacgagtcg gagttcgtcc ccgtcgccga gcagcagccc gacgccgaca      660
349 ccgagcagct cgtcgccgcc ccgtcgtcca acgacccgta catccagcgg ttccctcacga      720
351 tgtacaacaa gattcacgac ccagcgaacg gctacttcag cccgcaggga attccctacc      780
353 actcggtaga aacgctcatc gttgaggcac cggactacgg gcacgagaca acttcggagg      840
355 cgtacagctt ctggctctgg ctcgaagcga cgtacggcgc agtgaccggc aactggacgc      900
357 cgttcaacaa cgcttgagcg acgatggaac cgtacatgat cccgcagcac gcggaccagc      960
359 cgaacaacgc gtcgtacaac cccaacagcc cggcgtcgtg cgctccggaa gagccgctgc      1020
361 ccagcatgta cccggttgcc atcgacagca cgtgcccgtt tgggcacgac ccgctcggcg      1080
363 ccgaattgca gtcgacgtac ggcaactcgg acatttacgg catgcactgg ctggccgacg      1140
365 ttgacaacat ctacggatac ggcgacagcc cggcgggtgg ttgcgaactc ggtccttcgg      1200
367 ctaagggcgt ctctacatc aacacattcc agcgcggctc gcaggagtcc gtctggggaga      1260

```

## RAW SEQUENCE LISTING

DATE: 09/09/2005

PATENT APPLICATION: US/09/917,384B

TIME: 09:01:38

Input Set : A:\Sequence listing.txt

Output Set: N:\CRF4\09092005\I917384B.raw

```

369 cggtcaccca gccgacgtgc gacaacggca agtacggtgg ggcgcacggc tacgtcgacc 1320
371 tgttcatcca ggggttcgacg ccgccgcagt ggaagtacac cgatgccccg gacgccgacg 1380
373 cccgtgccgt ccaggctgcg tactgggcct acacctgggc atcggcgcag ggcaaggcaa 1440
375 gcgcgattgc cccgacgatc gccaaaggcg gccaaaccgg cgactacctg cggttactcg 1500
377 tctttgacaa gtacttcaag caggtcggca actgctacct ggccagctcc tgcctggag 1560
379 caaccggacg ccagagcgag acctacctga tcggctggta ctacgcctgg ggcggctcaa 1620
381 gccaaaggctg ggcctggcgc attggtgacg gcgccgcgca cttcggctac cagaatccgc 1680
383 ttgccgcgtg ggcgatgtcg aacgtgacac cgctcattcc gctctcgccc acggcaaaga 1740
385 gcgactgggc ggcgagcttg cagcgccagc tggagttcta ccagtggttg caatccgcgg 1800
387 aaggagccat tgcgggcggc gccaccaaca gctggaacgg caattacggg accccgcgg 1860
389 ccggagactc gaccttctac ggcatggcgt acgactggga gccggcttac cagcaccgcg 1920
391 cgagcaacaa ctggttcggc ttccaggcgt ggtccatgga acgggttgcc gagtactact 1980
393 acgtcacccg cgaccgaag gccaaaggcg tgctcgacaa gtgggtcgca tgggtgaagc 2040
395 cgaatgtcac caccggtgcc tcatggtcga ttccgtcgaa tttgtcctgg agcggccaac 2100
397 cggatacctg gaatccgagc aaccacggaa cgaatgccaa cctgcacgtg accatcacgt 2160
399 cgtccgggca ggacgtcggg gttgccgcgg cgctcgcgaa gacactcgag tactacgcgg 2220
401 caaatccgg cgatacggcc tcgcgcgacc tcgcgaaggg attgctcgac tccatgtgga 2280
403 acaacgacca ggacagctc ggtgtgagca caccggagac gcggaccgac tactctcggt 2340
405 tcaactcagg gtacgacccg acgactgggt acggcctcta catcccgtcg ggttggacgg 2400
407 ggaccatgcc caacggtgac caaatcaagc cgggtgcgac cttcctgagc atccggtcct 2460
409 ggtacaccaa ggatccgcag tggtcgaagg tgcaggcgta cctcaacggc gggcctgctc 2520
411 cgacgttcaa ctaccaccgg ttctgggcgg agtccgactt cgcgatggcg aacgccgatt 2580
413 ttggcatgct cttcccattc gggtcgcccc gcccgacccc gagcccgact ccgacgtcgt 2640
415 ccccgagccc gactccgagc agctcgccga cgccgtcgcc cagcccgtca ccgaccggcg 2700
417 acaccagccc gccgagcgtg ccgacgggtc ttcagggtcac cgggacaacg acgtcgtccg 2760
419 tgtcgctcag ctggaccgcg tccaccgaca acgtcggcgt cgcgcactac aacgtgtacc 2820
421 gaaacggcac gctggtgggt cagccgacag cgacgtcgtt cacggacacc ggcctggctg 2880
423 ctggcacgtc gtacagctac acagtggcgg ccgttgatgc ggccggtaac acgtcggcgc 2940
425 agagcttcgc cggtagacag gacgacggca tcgccgtcgc gagcccgtcg ccgagcccga 3000
427 ctccgagctc gtccccgagc ccaacggcgt cgccgacacc gtcaccgacg tccaccagcg 3060
429 gcgcatcgtg cactgctacc tacgttgtca atagcgactg gggtagcggc ttcacgacaa 3120
431 ccgtgaccgt gacgaacacc ggcaccaggg ccaccagtgg ctggacggtc acgtggagct 3180
433 ttgccggtaa tcagacggtc accaactact ggaacaccgc gctgacgcaa tccggaaagt 3240
435 cggtgaccgc aaagaacctg agttacaaca acgtcatcca acctggtcag tcgacgacct 3300
437 ttggattcaa cggaagtac tcaggaacaa acaccgcgcc gacgctcagc tgcacggcaa 3360
439 gctga
442 <210> SEQ ID NO: 3
443 <211> LENGTH: 34
444 <212> TYPE: PRT
445 <213> ORGANISM: Acidothermus cellulolyticus
448 <220> FEATURE:
449 <221> NAME/KEY: misc_feature
450 <223> OTHER INFORMATION: Potential signal peptide of Gux1
452 <400> SEQUENCE: 3
454 Met Pro Gly Leu Arg Arg Arg Leu Arg Ala Gly Ile Val Ser Ala Ala
455 1 5 10 15
458 Ala Leu Gly Ser Leu Val Ser Gly Leu Val Ala Val Ala Pro Val Ala
459 20 25 30
462 His Ala

```

RAW SEQUENCE LISTING ERROR SUMMARY      DATE: 09/09/2005  
PATENT APPLICATION:    US/09/917,384B      TIME: 09:01:39

Input Set : A:\Sequence listing.txt  
Output Set: N:\CRF4\09092005\I917384B.raw

Invalid <213> Response:

Use of "Artificial" only as "<213> Organism" response is incomplete,  
per 1.823(b) of New Sequence Rules. Valid response is Artificial Sequence.

Seq#:8

**VERIFICATION SUMMARY**

DATE: 09/09/2005

PATENT APPLICATION: US/09/917,384B

TIME: 09:01:39

Input Set : A:\Sequence listing.txt

Output Set: N:\CRF4\09092005\I917384B.raw